IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 5, No. 2, June 2016, pp. 64~71
ISSN: 2252-8938

An SVD based Real Coded Genetic Algorithm for Graph Clustering

Parthajit Roy*, J.K. Mandal**
* Department of Computer Science, the University of Burdwan
** Department of Computer Science & Engineering, the University of Kalyani

Article Info

Article history:

Received Mar 4, 2016
Revised May 8, 2016
Accepted May 25, 2016

Keyword:

Bipartite Graph
Cluster Analysis
Cluster Validity Index
Genetic Algorithm
Graph Clustering
Singular Value Decomposition

ABSTRACT

This paper proposes a novel graph clustering model based on genetic algorithm using a random point bipartite graph. The model uses random points distributed uniformly in the data space and the measurement of distance from these points to the test points have been considered as proximity. Random points and test points create an adjacency matrix. To create a similarity matrix, correlation coefficients are computed from the given bipartite graph. The eigenvectors of the singular value decomposition of the weighted similarity matrix are considered and the same are passed to an elitist GA model for identifying the cluster centers. The model has been tasted with the standard datasets and the performance has been compared with existing standard algorithms.

Corresponding Author:

Parthajit Roy,
Department of Computer Science,
The University of Burdwan India-713104
Email: roy.parthajit@gmail.com

1. INTRODUCTION

Clustering is grouping of objects that honors only the underlying proximity among the objects. Clustering is an unsupervised NP-hard grouping problem. There are several methods of clustering available in literature. The performance analysis of standard clustering models has been proposed by Xu et al [1] where as clustering and its underlying algebra is covered by Everitt et al [2].

Graph is a representation which models the objects and their relationship. The objects are generally represented as vertices or nodes of the graph and the relationships between two objects are represented using edges. Graph theory does not impose any restriction on the nature of the relationship. Any relation that exists in the real world can be incorporated as a relationship. The algebraic structures behind graphs are covered by Godsil and Royle [3].

Clustering on graph has gained enormous popularity now a days. In case of graph clustering, the data objects are represented in terms of vertices and the proximities are represented in terms of edges. i.e. two nodes are in proximity if there exists an edge between them. In case of weighted graph, the weight matrix is also considered for proximity measures. This means, the existence of an edge is not sufficient, the smaller the edge weight, the close the objects are. Once the problem has been casted to a graph model by representing the data points into objects and the distance between objects using the edges, the standard graph theoretic methods are used for clustering data points. A connectivity based graph clustering has been proposed by Hartuv and Shamir [4]. An web based application is made by He et al [5]. A survey on graph based clustering is done by Schaeffer [6].

The paper uses a bipartite graph based data clustering method where a bipartite graph has been created from the data points by taking some extra random points. The proximity of the sample points are

calculated based on their similarity towards the bipartite graph. A good literature is available on graph and the bipartite due to Bollob´as [7].

Singular value decomposition(SVD) is a factorization of matrices (not necessarily square) where a matrix is divided into three matrices of which two are orthogonal square matrices and the other is a diagonal matrix carries the eigenvalues. SVD is mainly used for dimensionality reduction. Some work on SVD based dimensionality reduction is done by Prabhu, and Anbazhagan [8]. SVD based document clustering has been considered by Dhillon [9]. Some earlier work on singular value based graph clustering is done by Drineas et al[10]. Genetic Algorithm is a modern heuristic searching tool which is applicable when the search space is huge and the exact solution of the problem is NP-hard.

Genetic algorithm often generates a pool of possible feasible solutions called the chromosomes and allows them to generate offspring solutions by mating. The offspring of good solutions are likely to be better. This assumption is the fundamental of genetic algorithm. The target of the genetic algorithm is to optimize an objective function to reach to a better solution. A novel elitist GA model is proposed by Deb et al [11]. The genetic algorithm based clustering technique are discussed by Maulik et al [12] but work does not consider graph based clustering; rather it considers fuzzy clustering on image data. Some more genetic algorithm based clustering is proposed by Che [13] whereas some application of clustering in genetic algorithm is due to Kala et al [14]. This paper proposes singular value decomposition based genetic algorithm model for clustering bipartite graph.

The paper is oriented as follows. Section 2. discusses the necessary mathematics used in the model. Section 3. proposes the model. The experimental setup has been covered in section 4. whereas the performance of the proposed model has been tasted and analyzed in section 5.. Conclusions are given in section 6. and the references come at the end.

2. THE MATHEMATICAL BACKGROUNDS

Several mathematical concepts are available for graph based clustering. A graph $G =< V, E >$ is an algebraic structure consists of ordered pair of the sets V and E where V is the set of vertices and the set E is a relation on $V \times V$. If $\forall v_i \in V, (v_i, v_i) \notin E$ and $\forall v_i, v_j \in V, (v_i, v_j) \in E \Leftrightarrow (v_j, v_i) \in E$, then the algebraic structure G is called an s-graph. If $\forall (v_j, v_j) \in E \Rightarrow (v_i, v_j)$ has a mapping to the real number set R, then the graph G is called a weighted graph.

A graph G is called a bipartite if the vertex set V can be decomposed into two disjoint sets V_1 and V_2 s.t. $\forall (v_i, v_j) \in E, v_i \in V_1 \ \& \ v_j \in V_2$. If the cardinality of $|V_1| = m$ and that of $|V_2| = n$, then the adjacency matrix of the bipartite graph G can be represented as given in equation 1.

$$G = \begin{bmatrix} \Theta_{|V_1|\times|V_1|} & B^T_{|V_2|\times|V_1|} \\ B_{|V_2|\times|V_1|} & \Theta_{|V_2|\times|V_2|} \end{bmatrix} \tag{1}$$

where $\Theta_{|V_1|\times|V_1|}$ and $\Theta_{|V_2|\times|V_2|}$ are $null$ matrices of respective orders, B is the relation from V_1 to V_2 and B^T is the transpose of B.

Singular value decomposition is a method that divides a rectangular matrix into three matrices as shown in equation 2 [15].

$$A_{n\times m} = U_{n\times n}.D_{n\times m}.V^T_{m\times m} \tag{2}$$

where where $UU^T = I_n$, U consists of orthonormal eigenvectors ot AA^T and V consists of the same of A^TA. D is a diagonal matrix that populates with the eigenvalues of U or V.

Genetic algorithm is a search based optimization technique for NP-hard problems. Genetic algorithm is inspired by the biological genetics where the chromosome of the parents decides the nature of their offspring. It is also believed that the good features of both the parents inherit to the offspring and the offspring may superior than its parents. The chromosome consists of several genes. Each gene is responsible for some property of the offspring. Properties like hair color, shape of nose etc are due to gene.

In genetic algorithm, the same concept has been inherited. Every individual component of the solution is encoded in a gene and chromosome is formed from the gene sets. Each chromosome is a candidate solution of the given problem. In genetic algorithm, a number of such solutions are created as initial population. The population is allowed for crossbreeding, called crossover. In such case, pairs of chromosome are chosen for mating. The result is the creation of new offspring. The offspring are then passed to a fitness function for their fitness or closeness towards the optimal solution. For this, an objective function is determined and the characteristics of the offspring are measured. Using the rule of survival of the fittest, the elitism model of the GA selects best n chromosome from the current pool and rest from the offspring. In this way the solution set evolves. During this evolution, some gene may get changed randomly to generate an entirely new solution. This is called mutation. After finite generations, an accepted solution is expected to occur [16].

The validity indexes have an important role in cluster analysis. It measures the quality of the results produced by the clustering algorithms. There are broadly two types of validity indexes, namely, internal and external. Almost all internal criteria considers the within cluster scatter and between cluster separation in someway or other. The present paper considers one of the internal cluster validity indexes, called Ball-Hall index [17] as the objective function for the genetic algorithm.

The external indexes, on the other hand, assume some prior knowledge about the sample distribution and measures the cluster produced accordingly. A good discussion on such indexes is given by Desgraupes [18] and Saha et al [19].

3. THE PROPOSED MODEL

The proposed technique is a bipartite graph based evolutionary model for clustering spatial non-categorical real valued data. Given the n data points in R^m dimension, the model considers it as a matrix of the form $A_{n\times m}$ of real numbers. Some random (let say p) points are generated in the sample space denoted as a matrix $S_{p\times m}$. A new graph $G =< V, E >$, s.t. $|V| = n + p$ is created in such a way that $\forall s_i S \& \forall a_k \in A$, an edge (s_i, a_k) is produced with the edge weight function as given in equation 3.

$$\delta(s_i, a_k) = e^{-\left(\sum_{j=1}^{m}(s_i[j]-a_k[j])^2\right)^{1/2}} \tag{3}$$

where $\delta(s_i, a_k) \in R$. This creates a bipartite graph whose adjacency matrix is given in equation 4.

$$Adj(G) = \begin{bmatrix} \Theta_{p\times p} & B^T_{n\times p} \\ B_{n\times p} & \Theta_{n\times n} \end{bmatrix} \tag{4}$$

Equation 4 is almost similar to Equation 1. The only difference is the values. In equation 1, the matrix B is populated with only 0 and 1 but in equation 4, matrix B is populated with real valued weight of the edges. From matrix of equation 4, matrix $B_{n\times p}$ is considered for computation. The singular value of the B matrix has been computed. i.e. B matrix has been decomposed into three matrices $B = U_{n\times n}.D_{n\times p}.V^T_{p\times p}$ as given in the equation 2. The process checks the eigenvalues of the diagonal entries of the matrix D and considers first $k + 1$ number of vectors from the matrix U where k is the number of clusters, a *priory*. These $k + 1$ vectors are of dimension n and they form a matrix $U_{n\times k+1}$. The model passes these vectors to the genetic algorithm to compute the clusters.

The genetic algorithm considered in the present model is an elitist model. In this model, every chromosome of a generation pool consists of fixed number of genes. The present paper assumes the number of clusters k as a prerequisite. Each chromosome $c_i = (g_1, g_2, g_3, \cdots, g_k)$ that of each gene $g_j = [r_1, r_2, r_3, \cdots, r_m]$, where m is the dimension of the matrix $U_{n\times k+1}$.

Initially the values of the genes have chosen as random values using equation 5.

$$\forall_{i\in k+1} r_i = rand(min(U_{n\times k+1}[i]), max(U_{n\times k+1}[i]) \tag{5}$$

This ensures that every gene will represent a valid point in the sample space and every chromosome is a valid solution to the problem. Initially a population of 20 chromosomes are considered. The process allows the chromosome for crossbreeding. It is assumed that the genes are unbreakable. During each crossover, a gene position is randomly selected and the genes of the corresponding positions are exchanged. To come out of the local optima, ξ percentage of chromosomes is allowed for mutation. If a chromosome is qualified for mutation, the random gene is selected from it and a random index of the gene is considered. The value of the index is populated with a random value which follows equation 5.

The model selects the chromosomes for the next generation using a elitist model. For this η percentage best chromosomes are selected from the parent pool and $100 - \eta$ percentage best chromosomes are selected from the next generation pool. They are allowed to pass for subsequent iterations.

The objective function of the present GA model is optimizing within cluster scatter, which is an internal validity criteria called Ball-Hall internal index. The formulation is given in equation 6 and 7.

$$w_i = \frac{1}{n_i}\sum_{j\in I^i}\left(U^{\{j\}}_{n\times k+1} - M^{\{j\}}\right)^2 \tag{6}$$

where I is an n dimensional vector. The value of $I^{\{i\}} = k_i; \forall i = \{1, 2, \cdots, k\}$, k being the number of clusters and $M^{\{i\}}$ follows g_i of equation 8, and the total dispersion is

$$\sigma = \frac{1}{k}\sum_{i=1}^{k} w_i \tag{7}$$

The objective is to maximize the value of σ of equation no 7. Every time, when pool of next generation is ready, the updating of the chromosomes c is done using the following rule (equation 8).

$$\forall g_i \in c, g_i = \frac{1}{n_i} \sum_{j \in I^{\{i\}}} U^{\{j\}}_{n \times (k+1)} \tag{8}$$

The genetic algorithm runs for maximum 100 iterations or till the updation is nominal. The final result is the best chromosome of the final iteration.

4. RESEARCH METHODOLOGY

To test the performance of the proposed model, the an experimental setup has been chosen. Three things are considered for this setup. These are the problem to consider, the dataset containing such problems, the measurement scales on which the outputs will be measured.

There are mainly two aspects that the clustering problems can be observed. One is on the basis of separability and the other is on the basis of the shape of the clusters. The problem of separability means whether the clusters are well separated or not. If the clusters are not well separated, the problem is touching problem. The shape of the clusters are of two types. Convex shaped and arbitrary shaped.

In figure 1(a), a convex and well separated sample has been shown whereas in figure 1(b), a sample is shown which is well separated but cluster shapes are non convex. In figure 1(c), the sample is a touching clusters problem. In sample of figure 1(c), the sample is also a problem of non-convex clusters.

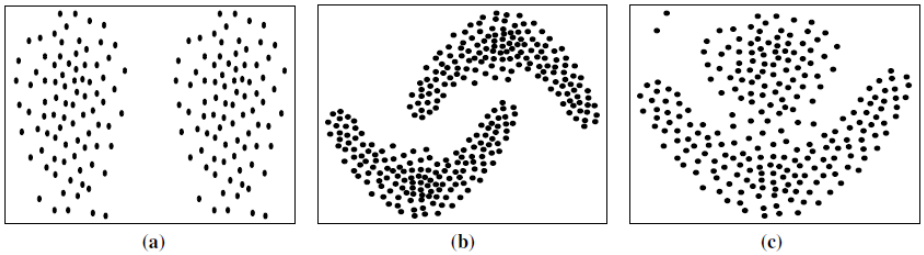

Figure 1. (a) The sample dataset is a compact convex shaped well separated dataset. (b) This sample shows a well separated dataset but the cluster shapes are not convex (The dataset has been generated by taking a part from the Flame [20] dataset). (c) This dataset is known as Flame [20] dataset. This dataset es an example of touching clusters. The shape is also non-convex for one cluster.

To test the performance of the model on the basis of discussion above, three datasets have been taken. iris [21],seeds [22] and Flame dataset. The iris dataset has four attributes, 150 instances and three clusters. One cluster is linearly separable from the other two whereas the other two clusters are not so. In Flame dataset, there are two attributes, 240 instances and two clusters. The clusters in the flame dataset are touching clusters and one cluster is convex shaped and the other is n non-convex shaped. The seeds dataset has 7 attributes, 210 instances and three clusters. The clusters are not linearly separable. As iris and seeds datasets are of higher dimensions, they are hard to visualize. The flame dataset is shown in figure 1(c).

The performance of the models have been measured in four different cluster validity indexes. These are Czekanowski-Dice, Folks-Mallows, Jaccard and Kulczynski. All indexes are external cluster validity indexes. The external indexes check the similarity between the output of the model with the actual distribution. The value if the external indexes lie between 0 and 1. If the similarity increases, the value goes up and becomes 1 if the result is identical to the actual.

There are other measures for checking the quality of the results. These are precision and F1 measure. Precision is the ratio of correctly identified points and total identified points. For example, if an algorithm produces n_1 number of points to belong to cluster C and out of which n_2 are actually the member of C, the precision is n_2/n_1. Recall or sensitivity on the other hand is the ratio of correctly identified points and actual points. Suppose, in the above example, if cluster C contains actually n number of points, then recall is n_2/n. The $F1$ measure is the harmonic mean of recall and precision which measures the accuracy of the cluster produced. Both precision and F-score produces value from 0 to 1 and 1 means perfect result in both the cases.

5. RESULTS AND DISCUSSIONS

To test the performance of the proposed technique, four standard models namely, k-means model, Hierarchical model (Single Linkage), Hierarchical model (Complete Linkage), Hierarchical model (Average Linkage) has been considered and the performances are shown in table 1, table 2 and table 3.

Table 1. Comparative study of the proposed model with that of other four standard models on Seeds [22] dataset. The performances have been measured on four external validity indexes.

External Index	Hierarchical (Complete Linkage)	Hierarchical (Single Linkage)	Hierarchical (Average Linkage)	K-Means Model	Proposed Model
Czekanowski Dice	0.7003818	0.4879907	0.8291612	0.8067365	0.7817273
Folks Mallows	0.7006848	0.554298	0.8291774	0.8067624	0.7817499
Jaccard	0.5389135	0.3227432	0.7081771	0.6760757	0.6416685
Kulczynki	0.700988	0.629615	0.8291935	0.8067883	0.7817725

From table 1 it can be seen the value of the indexes are higher for the proposed model than the complete linkage hierarchical model. This implies that the performance of the proposed model on seeds dataset is better than the complete linkage hierarchical model in all of the external indexes. The performance of the proposed model is better on seed dataset than the Single linkage Hierarchical also. The performance of the proposed model is, however, competitive on seeds dataset with that of average linkage Hierarchical model and standard k-means algorithm. The third, fourth and fifth column of table 1 shows this.

Table 2. Comparative study of the proposed model with that of other four standard models on Iris [21] dataset. The performances have been measured on four external validity indexes.

External Index	Hierarchical (Complete Linkage)	Hierarchical (Single Linkage)	Hierarchical (Average Linkage)	K-Means Model	Proposed Model
Czekanowski Dice	0.7671688	0.7414543	0.8404453	0.8206565	0.923432
Folks Mallows	0.7686371	0.7635171	0.8407289	0.820808	0.9234342
Jaccard	0.622282	0.5891358	0.7248000	0.6958588	0.8577554
Jaccard	0.7701083	0.7862363	0.8410127	0.8209596	0.9234363

The performance of the proposed model, on iris dataset, is better than complete linkage hierarchical model, single linkage hierarchical model, average linkage hierarchical model and standard k-means model in all external validity indexes considered. Table 2 shows the comparative studies respectively. It is clear from the table 2, that the index values for the proposed model is greater than that of all other models. This means that the distribution produced by the proposed model is more close the actual distribution than the distribution produced by the other algorithms.

Table 3. Comparative study of the proposed model with that of other four standard models on Flame [20] dataset. The performances have been measured on four external validity indexes.

External Index	Hierarchical (Complete Linkage)	Hierarchical (Single Linkage)	Hierarchical (Average Linkage)	K-Means Model	Proposed Model
Czekanowski Dice	0.6135203	0.6976339	0.7306216	0.7581325	0.8320555
Folks Mallows	0.6230364	0.7300235	0.73107	0.7586289	0.8342328
Jaccard	0.4425022	0.5356665	0.5755743	0.6104777	0.7124102
Jaccard	0.6327001	0.763917	0.7315187	0.7591256	0.8364157

The performance of the proposed model on Flame [20] dataset is also better in the external indexes. Table 3 shows this. The flame dataset is an example of touching cluster problems as well as non-convex cluster problems. In such cases also, the algorithm produces better result than the benchmark algorithms.

The accuracy of the proposed model on Seeds dataset has been shown in figure 2(a). The precision is competitive with the K-Means model and slightly less accurate than Hierarchical Model(Average Linkage). But it is more accurate than Complete Linkage and Single Linkage Hierarchical Modela. The same has been reflected in case of F1 score also. Figure 2(b) shows this.

The accuracy of the results produced by the proposed model is more accurate in case if Iris dataset. Figure 3(a) and figure 3(b) shows that the performance of the proposed model is considerably better than other benchmark models.

The performance of the proposed model is at per in case of precision with K-means algorithm on Flame dataset. The proposed model is better than other models in precision index. Figure 4(a) shows this. The F1 index, as shown in figure 4(b), is highest for the proposed model and hence the accuracy is also highest.

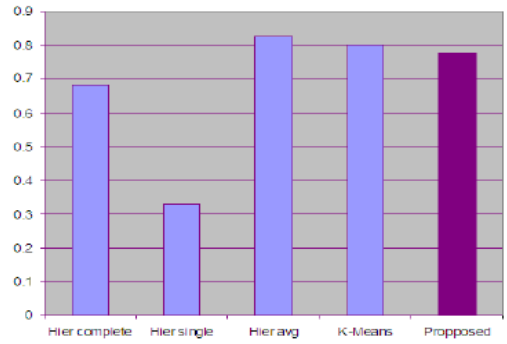

Figure 2a. Precision of the results produced by the five algorithms on Seeds dataset. The performance of the proposed model is shown in dark colour.

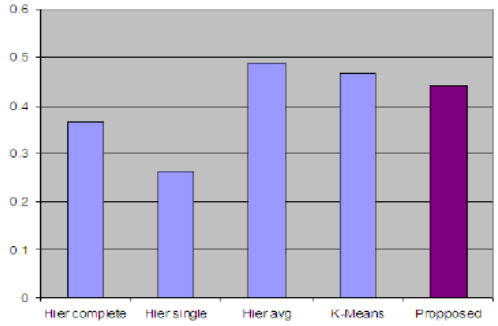

Figure 2b. F1 measure of the results produced by the five algorithms on Seeds dataset. The performance of the proposed model is shown in dark colour.

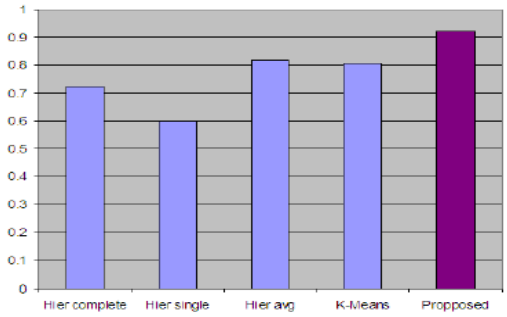

Figure 3a. Precision of the results produced by the five algorithms on Iris dataset. The performance of the proposed model is shown in dark colour.

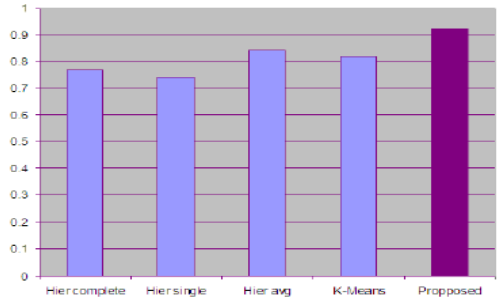

Figure 3b. F1 measure of the results produced by the five algorithms on Iris dataset. The performance of the proposed model is shown in dark colour.

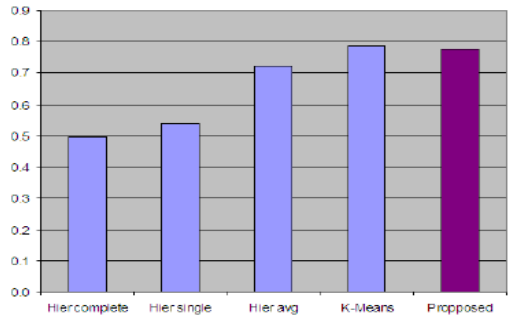

Figure 4a. Precision of the results produced by the five algorithms on Flame dataset. The performance of the proposed model is shown in dark colour.

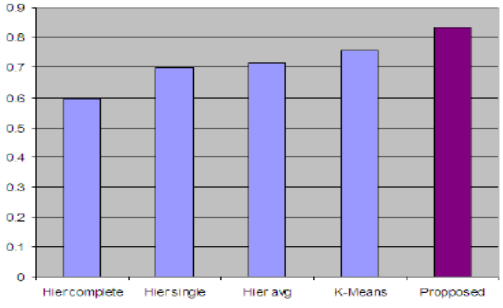

Figure 4b. F1 measure of the results produced by the five algorithms on Flame dataset. The performance of the proposed model is shown in dark colour.

6. CONCLUSION AND FUTURE SCOPE

This paper proposed a novel data clustering technique on singular matrix of bipartite graph based on genetic algorithm. The model shows good performance on standard datasets. The model is also advantageous because the elitist genetic algorithm model brigs the result out of the local optima. Though the model shows good result, the main attribute of the model is that it is computationally demanding. Further, the model can be improved by incorporating multi-objective evolutionary algorithms. Other soft-computing tools may also be used for further performance tuning.

ACKNOWLEDGMENT

The authors express the deep gratitude to the Department of Computer Science, the University of Burdwan, West Bengal, India and Department of Computer Science & Engineering, the University of Kalyani, West Bengal, India, for providing necessary infrastructure and support for the present work.

REFERENCES

[1] R Xu, D Wunsch-II. Survey of clustering algorithms. *IEEE Transactions on Neural Network*. 2005; 16(3): 645–678.
[2] BS Everitt, D Stahl, M Leese, S Landau. Cluster Analysis, 5th ed. John Wiley & Sons, 2011.
[3] C Godsil, G Royle. Algebraic Graph Theory, ser. Graduate Texts in Mathematics. *Springer*. 2001; 207.
[4] Hartuv, R Shamir. A clustering algorithm based on graph connectivity. Information Processing Letters, 2000.
[5] X He, CHD Hongyuan Zha, HD Simon. Web document clustering using hyperlink structures. Computational Statistics and Data Analysis. 2007; 41:19–45.
[6] SE Schaeffer. Graph clustering. *Elsevier Computer Science Review*. 2007; 1: 27–64.
[7] B Bollob´as. Modern Graph Theory. *Springer-Verlag*. 2005.
[8] P Prabhu, N Anbazhagan. Improving the performance of k-means clustering for high dimensional data set. *International Journal on Computer Science and Engineering (IJCSE)*. 2011; 3: 2317–2322.
[9] IS Dhillon. *Co-clustering documents and words using bipartite spectral graph partitioning*. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2001: 269–274. [Online]. Available: http://doi.acm.org/10.1145/502512.502550.
[10] P Drineas, A Frieze, R Kannan, S Vempala, V Vinay. Clustering large graphs via the singular value decomposition, Mach. Learn. 2004; 56(1-3): 9–33. [Online]. Available: http://dx.doi.org/10.1023/B:MACH.0000033113.59016.96.
[11] K Deb, A Pratap, S Agarwal, T. Meyarivan. Fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*. 2002; 6(2): 182–197.
[12] U Maulik, S Bandyopadhyay. Fuzzy partitioning using a real-coded variable-length genetic algorithm for pixel classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2003; 41(5): 1075–1081.
[13] Z Che. A hybrid algorithm for fuzzy clustering. *European Journal of Industrial Engineering*. 2012; 6: 50–67.
[14] R Kala, A Shukla, R Tiwari. Clustering-based hierarchical genetic algorithm for complex fitness landscapes. *International Journal of Intelligent Systems Technologies and Applications*, 2010; 9: 185–295.
[15] GH Golub, CF van Loan. Matrix Computations. Baltimore: Johns Hopkins University Press, 1996.
[16] K Deb, Multi-Objective Optimization using Evolutionary Algorithms. John Wiley Publication, 2001.
[17] GH Ball, DJ Hall. Isodata: A novel method of data analysis and pattern classifcation. Menlo Park: Stanford Research Institute. (NTIS No. AD 699616), 1965.
[18] B Desgraupes. Clustering indices. University Paris Ouest, Lab Modal‘X, Tech. Rep., April 2013.
[19] S Saha, S Bandyopadhyay. Performance evaluation of some symmetry-based cluster validity indexes. *IEEE Transactions on Systems*. Man, and Cybernetics: Part C:Applications and Reviews. 2009; 39(4): 420–425.
[20] L Fu, E Medico. Flame, a novel fuzzy clustering method for the analysis of dna microarray data. BMC bioinformatics, 2007.
[21] RA Fisher. Uci machine learning repository. 1936. [Online]. Available: http://archive.ics.uci.edu/ml
[22] M Charytanowicz, J Niewczas, P Kulczycki, PA Kowalski, S Lukasik, S Zak. Uci machine learning repository. 2010. [Online]. Available: http://archive.ics.uci.edu/ml.

BIOGRAPHIES OF AUTHORS

Parthajit Roy is a Master in Computer Applications from the University of Burdwan and currently pursuing his doctoral degree from the University of Kalyani. He is an Assistant Professor of the Department of Computer Science, the University of Burdwan. His area of interest is mainly Pattern Recognition, Computational Geometry and Graph Theory. His current research area is spectral graph theory based clustering.

J. K. Mandal is an M. Tech.(Computer Science, University of Calcutta), Ph.D.(Engg., Jadavpur University) in the field of Data Compression and Error Correction Techniques, Professor in Computer Science and Engineering, University of Kalyani, India. Life Member of Computer Society of India since 1992 and life member of cryptology Research Society of India. Former Dean, Faculty of Engineering, Technology & Management, working in the field of Network Security, Steganography, Remote Sensing & GIS Application, Image Processing and Pattern Recognition. 28 years of teaching and research experiences. Eleven Scholars awarded Ph.D. one submitted and 8 are pursuing. Total number of publications 355. Further info on his homepage: http://jkmandal.com